Nuovi modelli di IA: Explainable AI e Hybrid AI

Questo articolo è parte di una serie dedicata all'approfondimento legato all'IA Generativa, un campo in rapida evoluzione che sta ridefinendo i confini dell'innovazione tecnologica e della creatività computazionale.

L'intelligenza artificiale generativa sta permeando sempre più aspetti della nostra vita, plasmando il modo in cui interagiamo con il mondo e prendiamo decisioni. Dagli assistenti virtuali ai sistemi di supporto, dai veicoli a guida autonoma alle diagnosi mediche assistite, l'IA sta rivoluzionando interi settori e promettendo un futuro ricco di innovazione.

Tuttavia, questa diffusa adozione è accompagnata da una crescente preoccupazione connessa alla mancanza di trasparenza nel processo decisionale di molti sistemi di AI. Spesso definiti "scatole nere", questi sistemi producono risultati accurati, ma il percorso logico che li genera rimane oscuro e inaccessibile alla comprensione umana. Questa opacità solleva interrogativi cruciali sulla fiducia, la responsabilità e l'etica nell'utilizzo dell'AI.

In questo contesto, emerge l'importanza di modelli di Intelligenza Artificiale spiegabili e ibridi, un campo di ricerca in rapida evoluzione che mira a rendere i sistemi di AI più trasparenti e comprensibili. L'Explainable AI si propone di fornire spiegazioni chiare e interpretabili sul funzionamento interno di un modello di AI, consentendo agli utenti di comprendere come e perché un determinato risultato è stato raggiunto, mentre l'Hybrid AI prende il meglio dei diversi modelli di AI e li mette a fattore comune.

Opacità dell'IA: le sfide della Trasparenza Algoritmica

L'intelligenza artificiale che abbiamo conosciuto finora soffre di un grosso problema, l'opacità. Cosa significa opacità dell'AI? In ambito scientifico, l'opacità dell'AI si riferisce all'incapacità di comprendere il processo decisionale interno di un sistema di intelligenza artificiale, specialmente quelli basati su algoritmi complessi come il deep learning.

Ciò può comportare le seguenti conseguenze:

- **Mancanza di fiducia**: l'opacità può minare la fiducia dei progettisti e degli utilizzatori nell'accettazione dei risultati prodotti dall'AI, soprattutto in ambiti critici come la medicina, la fisica, la chimica, l'economia, dove la comprensione del processo decisionale è fondamentale per validare le scoperte,
- **Difficoltà di interpretazione**: l'opacità rende difficile l'interpretazione dei risultati e le previsioni generate dall'AI, limitando la capacità dei ricercatori ad estrarre dai dati nuove conoscenze e intuizioni;
- **Rischio di bias e discriminazione**: i modelli di AI opachi possono perpetuare bias presenti nei dati di training, portando a risultati discriminatori o ingiusti;
- **Problemi di debugging e miglioramento**: l'opacità rende difficile identificare e correggere errori o bias nei modelli di AI, ostacolando il loro miglioramento e l'affidabilità;
- **Ostacolo alla riproducibilità**: la mancanza di trasparenza nel processo decisionale dell'AI può rendere faticosa la riproduzione dei risultati e la validazione delle scoperte scientifiche.

L'opacità rappresenta una sfida importante anche per il metodo

scientifico, che notoriamente si basa sulla trasparenza, la riproducibilità e la comprensione dei fenomeni. L'utilizzo di modelli opachi può compromettere la validità delle scoperte scientifiche e la fiducia nella ricerca. Per affrontare questa criticità, la ricerca si è concentrate su queste tematiche:

- **Sviluppo di modelli di AI più interpretabili**: sono state sviluppate tecniche di modellazione, come l'Hybrid AI e l'Explainable AI, che mirano a creare sistemi più trasparenti e comprensibili;
- **Analisi e visualizzazione dei dati**: sono stati progettati strumenti per visualizzare e analizzare in dettaglio il comportamento dei modelli di AI, aiutando i ricercatori a comprendere il processo decisionale (debugging);
- **Validazione e verifica dei modelli di AI**: sono statti introdotti metodi per valutare l'accuratezza, l'affidabilità e l'equità dei modelli di AI, garantendo la validità dei risultati scientifici raggiunti.

In sintesi, l'opacità dell'AI rappresenta una sfida significativa per la ricerca. Superare questa opacità è fondamentale per garantire la fiducia, la trasparenza e la validità delle scoperte scientifiche nell'era dell'AI.

Explainable AI: aprire la Scatola Nera dell'Intelligenza Artificiale

L'intelligenza artificiale generativa (AI generativa) sta rapidamente trasformando il modo in cui creiamo e interagiamo con i contenuti digitali. Dai modelli linguistici di grandi dimensioni che generano testo realistico alle reti generative avversarie (GAN) che creano immagini e video sorprendenti, l'AI generativa sta aprendo nuove frontiere in diversi campi, dall'arte e l'intrattenimento alla scienza e alla medicina.

Tuttavia, il potere dell'AI generativa è accompagnato da una

crescente preoccupazione per la sua “scatola nera”. I modelli generativi spesso operano in modo opaco, rendendo difficile comprendere come arrivino a generare determinati output. Questa mancanza di trasparenza solleva diverse questioni cruciali:

- **Fiducia:** come possiamo fidarci dei risultati generati da un sistema di AI che non comprendiamo appieno?
- **Responsabilità:** chi è responsabile per gli output generati dall’AI, soprattutto se sono dannosi o imprecisi?
- **Controllo:** come possiamo controllare e guidare il processo creativo dell’AI generativa se non sappiamo come funziona?
- **Bias:** come possiamo identificare e mitigare i bias presenti nei modelli generativi se non siamo in grado di analizzarne il processo decisionale?

L’Explainable AI (XAI) si pone come un campo di ricerca cruciale nell’ambito dell’intelligenza artificiale, focalizzandosi sulla trasparenza e l’interpretabilità dei sistemi di AI. Mentre l’AI tradizionale mira principalmente a massimizzare le prestazioni predittive, l’XAI si concentra sulla comprensione del processo decisionale dei modelli, rendendoli accessibili e interpretabili dagli esseri umani.

Sebbene non esista una definizione univoca, l’XAI può essere concettualizzata come un insieme di principi, metodi e tecniche che mirano a:

- **Spiegare le decisioni:** XAI può fornire spiegazioni sui processi decisionali dei modelli generativi, aiutandoci a capire perché un modello ha generato un determinato output;
- **Identificare i bias:** XAI può aiutare a identificare i bias presenti nei dati di addestramento o nel modello stesso, consentendo di mitigare i loro effetti negativi;
- **Controllare il processo creativo:** XAI può fornire

strumenti per controllare e guidare il processo creativo dell'AI generativa, consentendo agli utenti di influenzare gli output generati;
- **Aumentare la fiducia:** Comprendere il funzionamento dei modelli generativi aumenta la fiducia negli output generati, promuovendone l'adozione in diversi ambiti.

Pertanto, l'Explainable AI è in grado di colmare una serie di criticità:

- **Superare l'opacità delle "black box"**: molti modelli di AI, specialmente quelli basati sul deep learning, operano come "scatole nere", rendendo difficile la comprensione del loro funzionamento interno. L'XAI mira ad "aprire" queste scatole nere, fornendo spiegazioni chiare e comprensibili sul processo decisionale;
- **Aumentare la fiducia e l'accettazione:** la trasparenza è fondamentale per costruire fiducia nei sistemi di AI e favorirne l'adozione in ambiti scientifici critici, dove la comprensione del processo decisionale è essenziale per validare le scoperte;
- **Garantire la responsabilità etica**: l'XAI consente di identificare e mitigare potenziali bias e discriminazioni nei modelli di AI, promuovendo un utilizzo responsabile ed etico nella ricerca;
- **Facilitare il miglioramento e il debugging**: comprendere il processo decisionale dei modelli di AI facilita il loro miglioramento, la correzione di errori e l'ottimizzazione delle prestazioni, elementi cruciali per la ricerca;
- **Promuovere la scoperta scientifica**: l'XAI può aiutare i ricercatori a estrarre nuove conoscenze e intuizioni dai dati, aprendo nuove prospettive di ricerca e accelerando il progresso scientifico.

Per raggiungere questi obiettivi, è necessario che l'XAI si avvalga di un set di tecniche tra cui:

- **Modelli intrinsecamente interpretabili**: è fondamentale sviluppare modelli per la progettazione più semplici e trasparenti, come gli alberi decisionali, i modelli lineari e le regole di associazione;
- **Spiegazioni post-hoc**: occorre effettuare l'analisi del comportamento di un modello di AI dopo il suo addestramento, utilizzando tecniche come:
    - **Analisi di sensitività**: in grado di valutare l'influenza di ciascuna variabile di input sul risultato finale;
    - **Visualizzazione dei dati**: occorre utilizzare grafici e mappe per visualizzare il processo decisionale;
    - **LIME (Local Interpretable Model-Agnostic Explanations)**: è opportuno approssimare localmente il comportamento di un modello complesso con un modello più semplice;
    - **SHAP (SHapley Additive exPlanations)**: è possibile spiegare l'output di un modello in termini di contributo di ciascuna variabile di input;
- **Metodi di attenzione**: è necessario identificare le parti più importanti dei dati di input che influenzano la decisione del modello, fornendo informazioni su quali aspetti dei dati sono stati considerati più rilevanti;
- **Spiegazioni controfattuali**: è utile fornire informazioni su come modificare gli input per ottenere un output desiderato, aiutando a comprendere il processo decisionale del modello e a identificare potenziali interventi.

Nonostante il campo di ricerca avente ad oggetto l'XAI stia progredendo significativamente, occorre affrontare ulteriori sfide:

- **Definire il concetto di "spiegabilità"**, ovvero stabilire i criteri oggettivi e condivisi per valutare la qualità e l'efficacia di una spiegazione, adattandoli ai diversi

contesti scientifici;

- **Estendere la spiegabilità per diversi utenti**, ovvero adattare le spiegazioni alle esigenze e alle competenze di diversi tipi di utenti, dai ricercatori agli esperti di dominio;
- **Ricercare nuove tecniche di XAI** per continuare a sviluppare nuove tecniche per spiegare modelli di AI sempre più complessi, garantendo la loro applicabilità in diversi ambiti scientifici;
- **Integrare l'XAI nel ciclo di vita dell'AI** per incorporare l'XAI nella progettazione, lo sviluppo e l'implementazione dei sistemi di AI, promuovendo la trasparenza e la responsabilità fin dalle prime fasi della ricerca.

Ecco alcuni esempi di XAI per l'AI generativa:

- **Visualizzazione delle attivazioni:** la visualizzazione delle attivazioni dei neuroni in una rete neurale può aiutare a capire quali parti del modello sono responsabili per la generazione di specifici output;
- **Analisi di sensibilità:** l'analisi di come le variazioni negli input influenzano gli output generati può fornire informazioni sul funzionamento del modello,
- **Generazione di spiegazioni testuali:** lo sviluppo di modelli in grado di generare spiegazioni testuali per i loro output, rendendoli più comprensibili agli utenti.

XAI è un campo in rapida evoluzione con il potenziale per rivoluzionare il modo in cui interagiamo con l'AI generativa. Applicando XAI, possiamo rendere i modelli generativi più trasparenti, responsabili e affidabili, aprendo la strada a un utilizzo più etico e sicuro di questa potente tecnologia. L'integrazione di XAI nell'AI generativa non solo favorirà la fiducia e l'accettazione di questa tecnologia, ma stimolerà anche nuove forme di creatività e innovazione.

Hybrid AI: integrare Reti Neurali e Ragionamento Simbolico

L’intelligenza artificiale ibrida (Hybrid AI), che combina la potenza delle reti neurali con la capacità di ragionamento simbolico, si presenta come un potente catalizzatore per l’evoluzione dell’AI generativa. Questo connubio apre nuove frontiere per la creatività, l’intelligenza e l’affidabilità dei sistemi di AI, superando i limiti dei singoli approcci.

I modelli generativi, pur producendo risultati sorprendenti, spesso mancano di trasparenza nel processo decisionale (vd. paragrafo precedente). L’Hybrid AI, integrando il ragionamento simbolico, può fornire spiegazioni più chiare e comprensibili sul perché e come un modello generativo produce un determinato output. Questo aumenta la fiducia negli utenti e consente un controllo più preciso sul processo creativo.

L’Hybrid AI si basa su questi principi:

- **Complementarietà delle tecniche**: è notorio che vi siano diverse tecniche di modellazione che eccellono in compiti diversi. Ad esempio, l’apprendimento automatico eccelle nell’individuare pattern e fare previsioni basate su grandi quantità di dati, ma può essere opaco e difficile da interpretare. Il ragionamento simbolico, d’altra parte, offre trasparenza e capacità di ragionamento logico, ma può avere difficoltà a gestire l’incertezza e la complessità del mondo reale. L’Hybrid AI mira a sfruttare il meglio di entrambi i mondi, combinando la capacità di apprendimento automatico di estrarre informazioni dai dati con la capacità del ragionamento simbolico di elaborare conoscenze e regole;
- **Sinergia e integrazione**: l’Hybrid AI non si limita a un semplice giustapposizione di tecniche, ma mira a una loro integrazione profonda, in cui le diverse componenti interagiscono e si completano a vicenda, creando un

sistema olistico con capacità superiori alla somma delle sue parti;
- **Modellazione della cognizione umana**: l'Hybrid AI si ispira spesso alla cognizione umana, che combina intuizione, ragionamento e apprendimento. Integrando diverse tecniche di AI, si cerca di emulare la flessibilità e l'adattabilità dell'intelligenza umana.

Esistono diverse modalità per combinare differenti tecniche di modellazione in un sistema ibrido, i principali metodi di approccio possono essere individuati nei seguenti:

- **Integrazione dei modelli ad apprendimento automatico con sistemi basati sulle regole**: l'apprendimento automatico può essere utilizzato per estrarre le regole dai dati, che poi vengono integrate in un sistema basato sul ragionamento simbolico, consentendo al sistema di apprendere da grandi quantità di dati e di utilizzare la conoscenza simbolica per ragionare e prendere decisioni;
- **Combinazione delle reti neurali con algoritmi evolutivi**: gli algoritmi evolutivi possono essere utilizzati per ottimizzare la struttura e i parametri delle reti neurali, migliorando la loro efficienza e la loro capacità di generalizzazione;
- **Utilizzo di tecniche di fuzzy logic per gestire l'incertezza**: la fuzzy logic consente di rappresentare e di effettuare ragionamenti su concetti vaghi e incerti, migliorando la capacità dei sistemi di AI nella gestione della complessità del mondo reale e di prendere decisioni in situazioni ambigue;
- **Integrazione dei modelli di deep learning con i metodi di ragionamento basati sulla logica**: combinare la potenza della rappresentazione del deep learning con la capacità di ragionamento e l'inferenza della logica, consentendo ai sistemi di AI di apprendere da grandi quantità di dati e di utilizzare la logica per ragionare e spiegare le proprie decisioni.

Tutto ciò può portare dei vantaggi significativi in termini di:

- **Robustezza e affidabilità**: i sistemi ibridi sono meno suscettibili a errori e bias rispetto ai sistemi basati su un singolo approccio, in quanto le diverse componenti possono compensare le debolezze reciproche;
- **Flessibilità e adattabilità**: i sistemi ibridi possono adattarsi più facilmente a nuovi contesti e compiti, grazie alla loro capacità di combinare diverse strategie di apprendimento e di ragionamento;
- **Interpretabilità e trasparenza**: l'Hybrid AI può favorire lo sviluppo di sistemi di AI più trasparenti e comprensibili, in quanto l'integrazione di tecniche come il ragionamento simbolico può rendere il processo decisionale più esplicito;
- **Efficienza**: l'Hybrid AI può ottimizzare l'utilizzo delle risorse e migliorare le prestazioni dei sistemi di AI, combinando le diverse tecniche in modo sinergico.

Nonostante il suo potenziale, l'Hybrid AI deve affrontare diverse sfide:

- **Sviluppo di framework e strumenti specifici**: sono necessari nuovi strumenti e framework per facilitare l'integrazione di diverse tecniche di AI in modo efficiente e scalabile;
- **Gestione della complessità**: i sistemi ibridi possono essere più complessi da progettare, implementare e gestire rispetto ai sistemi basati su un singolo approccio. Pertanto, sono necessarie nuove metodologie e tecniche per affrontare la complessità dei sistemi ibridi;
- **Valutazione e benchmarking**: è necessario sviluppare nuove metriche e metodi di valutazione per confrontare e valutare le prestazioni dei sistemi di AI ibrida in modo oggettivo e rigoroso.

L’Hybrid AI, combinando l’AI generativa con l’AI simbolica, offre diverse opportunità:

- **Spiegabilità:** l’AI simbolica, basata su regole e logica, introduce trasparenza nel processo generativo. Le regole possono essere utilizzate per spiegare le decisioni prese dal sistema e per giustificare gli output generati;
- **Ragionamento e Generalizzazione:** l’AI simbolica consente di integrare conoscenza pregressa e di ragionare su di essa, migliorando la capacità di generalizzazione dell’AI generativa e consentendole di affrontare situazioni nuove in modo più efficace;
- **Controllo e Bias Mitigation:** l’AI simbolica può essere utilizzata per definire vincoli e regole che guidano il processo generativo, limitando i bias e garantendo che gli output siano conformi a determinati criteri o standard etici.

Ecco alcune applicazioni Concrete dell’Hybrid AI per l’AI Generativa:

- **Generazione di Contenuti Controllata:** nell’ambito della scrittura creativa, l’Hybrid AI può garantire che il testo generato rispetti determinate regole grammaticali, stilistiche o di contenuto, aumentando la coerenza e la qualità del testo;
- **Creazione di Arte e Musica:** l’Hybrid AI può combinare la creatività dell’AI generativa con la struttura e l’armonia fornite dall’AI simbolica, consentendo la creazione di opere d’arte e musica più sofisticate e innovative;
- **Design e Ingegneria:** l’Hybrid AI può generare progetti che soddisfano specifici requisiti funzionali ed estetici, combinando l’esplorazione creativa con la validazione basata su regole e vincoli;
- **Scoperta Scientifica:** l’Hybrid AI può accelerare la scoperta di nuovi materiali, farmaci o soluzioni

ingegneristiche, generando ipotesi e modelli che vengono poi validati e raffinati attraverso il ragionamento simbolico.

L'Hybrid AI promette di rivoluzionare l'AI generativa, aprendo nuove frontiere per la creatività, l'innovazione e la risoluzione di problemi complessi. La combinazione di intuizione e ragione, di apprendimento e logica, rappresenta un passo fondamentale verso lo sviluppo di un'AI più potente, versatile e affidabile.

Uno sguardo al futuro

Gli sviluppi dell'intelligenza artificiale sono stati così significativi in termini sia di dimensione, che di velocità, che non c'è stato sempre né il tempo, né l'opportunità, di comprendere pienamente la loro sicurezza e la loro effettiva rilevanza.

Di seguito sono esposite alcune branche dell'Intelligenza Artificiale, in evoluzione e in grado di rispondere alle nuove richieste sempre più incombenti, e il loro apporto al miglioramento dell'Intelligenza Artificiale Generativa.

Advanced Machine Learning

L'Intelligenza Artificiale Generativa (AI Generativa) sta vivendo una fase di rapida evoluzione alimentata anche dai progressi nell'Advanced Machine Learning (ML avanzato). Questo connubio apre nuove frontiere e opportunità, consentendo la creazione di modelli generativi sempre più sofisticati e capaci. In tal senso, l'apprendimento automatico (*Machine Learning*) ha già rivoluzionato il campo dell'intelligenza artificiale, consentendo ai computer di apprendere dai dati senza essere esplicitamente programmati. In tal senso, ha permesso di ottenere risultati straordinari in diversi ambiti, come il riconoscimento di immagini, l'elaborazione del linguaggio naturale e la previsione di serie temporali.

Tuttavia, l’apprendimento automatico tradizionale presenta alcune limitazioni che ne ostacolano il pieno potenziale:

- **Dipendenza dai dati etichettati**: molti algoritmi di apprendimento automatico richiedono grandi quantità di dati etichettati, la cui acquisizione può essere costosa e laboriosa;
- **Generalizzazione limitata**: i modelli di apprendimento automatico possono avere difficoltà a generalizzare nuovi dati o contesti diversi da quelli su cui sono stati addestrati;
- **Mancanza di interpretabilità**: spesso è difficile comprendere il processo decisionale dei modelli di apprendimento automatico, il che può limitarne l’affidabilità e l’accettazione.

L’Apprendimento Automatico Avanzato (*Advanced Machine Learning*) si propone di superare queste limitazioni, esplorando nuove frontiere e aprendo nuove possibilità per la ricerca e l’innovazione tecnologica. In particolare, può affrontare alcune sfide che limitano l’AI generativa, come:

- **Qualità e diversità:** è possibile migliorare la qualità e la diversità dei contenuti generati, riducendo artefatti e ripetizioni;
- **Controllo e personalizzazione:** è in grado di aumentare il controllo sui contenuti generati, consentendo la personalizzazione in base alle esigenze degli utenti;
- **Generalizzazione e creatività:** può sviluppare modelli in grado di generalizzare a nuovi contesti e di esprimere maggiore creatività;
- **Spiegabilità e affidabilità:** può rendere i modelli generativi più trasparenti e affidabili, fornendo spiegazioni sul loro funzionamento.

Per superare queste limitazioni, l’apprendimento automatico avanzato esplora nuove frontiere:

- **Apprendimento auto-supervisionato (SSL)**: questa tecnica consente ai modelli di apprendere da soli da grandi quantità di dati non strutturati, riducendo la dipendenza da dati etichettati. L'obiettivo è quello di estrarre automaticamente informazioni significative dai dati, scoprendo pattern e relazioni nascoste. SSL può migliorare l'efficienza dell'addestramento e la generalizzazione dei modelli generativi. Alcuni esempi di apprendimento auto-supervisionato includono:
    - **Pre-addestramento di modelli linguistici**: modelli come BERT e GPT-3 sono pre-addestrati su enormi quantità di testo, imparando a prevedere parole mancanti o a generare testo coerente;
    - **Apprendimento di rappresentazioni visive**: modelli come SimCLR e MoCo imparano a generare rappresentazioni visive significative da immagini non etichettate;
- **Apprendimento federato**: questa tecnica consente di addestrare modelli di AI su dati distribuiti su diversi dispositivi, preservando la privacy e la sicurezza dei dati. L'apprendimento federato è particolarmente utile in ambiti come la medicina e la finanza, dove i dati sono spesso sensibili e decentralizzati. Un esempio di applicazione è l'addestramento di modelli di diagnosi medica su dati provenienti da diversi ospedali, senza la necessità di condividere i dati dei pazienti (vd. Apposito paragrafo);
- **Meta-apprendimento**: questa tecnica si concentra sullo sviluppo di modelli di AI in grado di "apprendere come apprendere", adattandosi rapidamente a nuovi compiti e contesti. L'obiettivo è quello di creare sistemi di AI più flessibili e versatili, in grado di generalizzare a nuovi problemi e di apprendere in modo continuo. Un esempio di meta-apprendimento è l'algoritmo MAML (Model-Agnostic Meta-Learning), che apprende un insieme di parametri iniziali che consentono al modello di adattarsi rapidamente a nuovi compiti. Il meta-

apprendimento può rendere i modelli generativi più flessibili e versatili;

- **Apprendimento per rinforzo (RL)**: questa tecnica consente ai modelli di apprendere interagendo con un ambiente e ricevendo feedback sotto forma di ricompense o penalità. L'apprendimento per rinforzo è particolarmente utile per lo sviluppo di agenti intelligenti, come robot autonomi e sistemi di controllo, In AI generativa, RL può essere utilizzato per ottimizzare la generazione di contenuti in base a obiettivi specifici,
- **Apprendimento profondo spiegabile**: questa area di ricerca si concentra sullo sviluppo di modelli di deep learning più interpretabili, consentendo di comprendere il processo decisionale e di aumentare la fiducia negli output del modello;
- **Modelli probabilistici profond (Deep Probabilistic Models)**: questa tecnica combina le reti neurali con modelli probabilistici per rappresentare l'incertezza e generare contenuti con maggiore varietà e controllo;
- **AI neuro-simbolica**: questo metodo integra le reti neurali con sistemi di ragionamento simbolico per creare modelli generativi più spiegabili e capaci di ragionamento astratto.

Nonostante i significativi progressi effettuati, l'AML deve ancora affrontare diverse sfide:

- **Scalabilità**: è fondamentale sviluppare algoritmi in grado di gestire dataset sempre più grandi e complessi;
- **Robustezza**: è necessario creare modelli di AI resistenti a errori e perturbazioni nei dati;
- **Etica e responsabilità**: è essenziale garantire che l'apprendimento automatico avanzato sia utilizzato in modo responsabile ed etico, evitando bias e discriminazioni.

L'ML avanzato apre nuove opportunità per l'AI generativa in diversi ambiti:

- **Creazione di contenuti realistici**: generare immagini, video, audio e testo di alta qualità, indistinguibili da quelli reali;
- **Scoperta scientifica**: generare nuove molecole, materiali e prodotti con proprietà desiderate;
- **Arte e creatività**: assistere artisti e creativi nella generazione di nuove opere d'arte, musica e design;
- **Personalizzazione dell'esperienza utente**: creare esperienze personalizzate in base alle preferenze individuali degli utenti.

L'ML avanzato sta trasformando l'AI generativa, consentendo la creazione di modelli sempre più potenti e versatili. Questa sinergia apre nuove frontiere in diversi ambiti, ma richiede anche un approccio responsabile per garantire un uso etico e benefico di questa tecnologia.

Embodied AI: un corpo per l'IA Generativa

L'Intelligenza Artificiale Generativa (Generative AI) consente di generare testi, immagini e codice con sorprendente creatività. Tuttavia, questi modelli operano in un mondo astratto, privi di un corpo fisico che permetta loro di interagire direttamente con l'ambiente. È qui che entra in gioco l'Embodied AI, un campo di ricerca che esplora l'integrazione dell'AI in corpi robotici, aprendo nuove e affascinanti prospettive per l'AI generativa.

L'Intelligenza Artificiale Embodied (Embodied AI) rappresenta un cambio di paradigma nello studio dell'AI, spostando il focus dalla pura elaborazione di informazioni all'interazione fisica con il mondo. Invece di limitarsi ad elaborare informazioni in modo digitale, l'Embodied AI si concentra sull'integrazione dell'intelligenza in corpi fisici, come robot o agenti virtuali, che interagiscono con l'ambiente circostante[1].

Tradizionalmente, l'AI si è concentrata sullo sviluppo di algoritmi e modelli in grado di elaborare informazioni e risolvere problemi in modo astratto, senza una connessione diretta con il mondo fisico. L'Embodied AI, al contrario, enfatizza l'importanza dell'incarnazione dell'intelligenza in corpi fisici. Questo approccio si basa sull'idea che la cognizione sia strettamente legata all'esperienza fisica e all'interazione sensomotoria[2].

L'Embodied AI si fonda su una serie di teorie e principi:

- **Cognizione incorporata**: l'Embodied AI si basa sulla teoria della cognizione incorporata, che sostiene che la cognizione è strettamente legata all'esperienza fisica e all'interazione con l'ambiente. Secondo questa teoria, il corpo non è un semplice strumento dell'intelligenza, ma un elemento costitutivo della cognizione stessa;
- **Interazione sensomotoria**: l'Embodied AI enfatizza l'importanza dell'interazione sensomotoria, ovvero la capacità di percepire l'ambiente attraverso i sensori e di agire su di esso attraverso gli attuatori. Questa interazione dinamica con il mondo consente all'Embodied AI di apprendere dall'esperienza e di adattarsi a nuove situazioni;
- **Apprendimento situato**: l'Embodied AI si concentra sull'apprendimento situato, ovvero l'apprendimento che avviene in un contesto specifico e che è guidato dall'interazione con l'ambiente. Questo tipo di apprendimento consente all'AI Embodied di acquisire conoscenze e competenze rilevanti per il suo ambiente e i suoi obiettivi.

Vediamo quali possono essere le implicazioni dell'Embodied AI:

- **Robotica cognitiva**: la robotica cognitiva si concentra sullo sviluppo di robot autonomi in grado di interagire con l'ambiente, apprendere dall'esperienza e risolvere problemi in modo flessibile e adattabile[3];

- **Agenti virtuali embodied**: gli agenti virtuali embodied sono programmi che simulano il comportamento di agenti intelligenti in ambienti virtuali. Questi agenti possono essere utilizzati per studiare l'intelligenza artificiale, la cognizione umana e l'interazione sociale;
- **Interfacce cervello-computer**: le interfacce cervello-computer consentono di collegare il cervello umano a dispositivi esterni, come protesi robotiche o computer. Questa tecnologia può essere utilizzata per sviluppare nuove forme di Embodied AI che integrano l'intelligenza umana con quella artificiale.

I principali vantaggi connessi allo sviluppo dell'Embodied AI possono essere sintetizzati in:

- **Comprensione della cognizione**: l'Embodied AI può fornire nuove intuizioni sul funzionamento della cognizione umana, studiando come l'intelligenza emerge dall'interazione tra corpo, cervello e ambiente;
- **Sviluppo di robot autonomi**: l'Embodied AI può consentire lo sviluppo di robot autonomi più intelligenti e adattabili, in grado di operare in ambienti complessi e di collaborare con gli esseri umani;
- **Nuove interfacce uomo-macchina**: l'Embodied AI può portare allo sviluppo di nuove interfacce uomo-macchina più intuitive e naturali, che consentono una migliore interazione tra esseri umani e computer;
- **Modellazione di sistemi complessi**: l'Embodied AI può essere utilizzata per modellare e simulare sistemi complessi, come ecosistemi, società e organismi biologici, fornendo nuove prospettive di ricerca.

I modelli di Generative AI attuali, pur sofisticati, mancano di una comprensione profonda del mondo fisico. L'Embodied AI, invece, permette all'AI di "vivere" il mondo attraverso un corpo robotico, acquisendo informazioni sensoriali e interagendo con l'ambiente. Questa esperienza incarnata può

arricchire l’AI generativa in diversi modi:

- **Comprensione contestuale:** l’interazione con l’ambiente fornisce un contesto cruciale per la generazione di contenuti. Un robot che “vede” una mela può generare descrizioni più accurate e creative rispetto a un modello che ha appreso solo da immagini;
- **Apprendimento grounded:** l’Embodied AI permette di “radicare” la conoscenza in esperienze concrete, creando rappresentazioni più robuste e significative;
- **Creatività situata:** la creatività non è un processo astratto, ma è influenzata dal contesto e dalle interazioni. Un robot che “vive” in un ambiente può sviluppare una creatività più ricca e sfaccettata.

L’Embodied AI può spingere l’AI generativa oltre la semplice produzione di contenuti digitali, consentendo la creazione di opere fisiche e l’interazione con il mondo reale:

- **Arte robotica:** Robot dotati di AI generativa potrebbero dipingere, scolpire o comporre musica, esprimendo la loro creatività in forme nuove;
- **Design generativo:** L’Embodied AI può guidare la progettazione e la fabbricazione di oggetti fisici, adattandoli alle esigenze dell’ambiente e degli utenti;
- **Interazione uomo-robot:** Robot con capacità generative potrebbero comunicare e collaborare con gli esseri umani in modi più naturali e intuitivi.

L’integrazione di Embodied AI e Generative AI presenta sfide importanti:

- **Complessità:** sviluppare robot con capacità sensoriali, motorie e cognitive avanzate è un compito complesso;
- **Generalizzazione:** i modelli generativi devono essere in grado di generalizzare a diversi ambienti e situazioni;
- **Etica:** è fondamentale garantire che l’Embodied AI sia sviluppata e utilizzata in modo responsabile ed etico.

Nonostante le sfide, l'Embodied AI offre un potenziale enorme per l'AI generativa. In futuro, potremmo vedere robot che non solo generano contenuti creativi, ma che interagiscono con il mondo in modi significativi, contribuendo a risolvere problemi e a migliorare la nostra vita.

In conclusione, l'Embodied AI rappresenta una nuova frontiera per l'AI generativa, offrendo la possibilità di creare sistemi di AI più intelligenti, creativi e integrati con il mondo fisico.

Generalized Artificial Intelligence

L'attuale generazione di modelli di AI generativa presenta delle limitazioni, tra cui la difficoltà nel comprendere il contesto, la mancanza di senso comune e la tendenza a generare output incoerenti o irrealistici. L'avvento dell'Intelligenza Artificiale Generale (AGI), ovvero un'AI con capacità cognitive simili a quelle umane, potrebbe rivoluzionare l'AI generativa, aprendo nuove frontiere di creatività e innovazione.

L'Intelligenza Artificiale Generalizzata (AGI), spesso definita anche come "IA forte" o "IA completa", rappresenta un obiettivo ambizioso nel campo dell'intelligenza artificiale. A differenza dell'AI ristretta, che eccelle in compiti specifici, l'AGI mira a replicare l'intelligenza umana nella sua totalità, creando macchine in grado di apprendere, ragionare, risolvere problemi e interagire con il mondo in modo flessibile e adattabile, proprio come gli esseri umani[4] [5].

L'AGI si distingue dall'AI ristretta per alcune caratteristiche fondamentali:

- **Generalità**: la capacità di apprendere e svolgere un'ampia gamma di compiti, anziché essere limitata a un dominio specifico;

- **Adattabilità**: la possibilità di trasferire conoscenze e competenze da un dominio all'altro e di adattarsi a nuove situazioni e ambienti;
- **Astrazione**: l'opportunità di ragionare a livelli di astrazione elevati, di formare concetti e di generalizzare da esempi specifici;
- **Coscienza e autoconsapevolezza**: sebbene ancora oggetto di dibattito, alcuni ricercatori ritengono che l'AGI possa sviluppare una forma di coscienza e autoconsapevolezza, simile a quella umana.

Al contempo, lo sviluppo concreto di questa modellazione presuppone la risoluzione di una serie di criticità:

- **Complessità del cervello umano**: il cervello umano è un sistema incredibilmente complesso, con miliardi di neuroni interconnessi. Replicarne il funzionamento richiede una comprensione profonda dei meccanismi neurali e cognitivi;
- **Mancanza di una teoria unificata dell'intelligenza**: ancora non esiste una teoria completa e condivisa su cosa sia l'intelligenza e come funzioni;
- **Rappresentazione della conoscenza**: la rappresentazione della conoscenza del mondo, in modo che possa essere utilizzata da un sistema di AGI per ragionare e risolvere problemi;
- **Apprendimento continuo**: come sviluppare sistemi di AGI in grado di apprendere in modo continuo e autonomo, adattandosi a nuove informazioni e situazioni;
- **Motivazione e obiettivi**: come dotare i sistemi di AGI di motivazioni e obiettivi intrinseci, in modo che possano agire in modo autonomo e finalizzato.

Per affrontare queste sfide si utilizzano i seguenti approcci scientifici in parte esaminati:

- **AI Simbolica**: si basa sulla manipolazione di simboli e sulla logica formale per rappresentare la conoscenza e

il ragionamento,

- **Reti neurali artificiali**: simulano il funzionamento del cervello umano attraverso reti di neuroni artificiali interconnessi,
- **Approcci evolutivi**: utilizzano algoritmi genetici per far evolvere sistemi di AI con capacità cognitive sempre più avanzate,
- **AI Ibrida**: combina diverse tecniche di AI per sfruttare i punti di forza di ciascun approccio,
- **Cognizione incorporata**: studia l'intelligenza artificiale in relazione all'interazione con un corpo fisico e un ambiente.

Ecco alcuni modi in cui l'AGI potrebbe potenziare l'AI generativa:

1. **Comprensione del Contesto e del Senso Comune:** l'AGI, con la sua capacità di comprendere il mondo in modo olistico e di applicare il senso comune, potrebbe consentire all'AI generativa di:

   - **Generare contenuti più coerenti e pertinenti al contesto:** potrebbe aiutare l'AI generativa a interpretare le sfumature del linguaggio, a comprendere le relazioni tra i concetti e a generare output che siano logicamente validi e coerenti con il contesto,
   - **Creare contenuti più originali e creativi:** potrebbe fornire all'AI generativa la capacità di combinare idee in modi nuovi e inaspettati, di esplorare diverse prospettive e di generare contenuti veramente innovativi,
   - **Adattare i contenuti a specifici pubblici e obiettivi:** potrebbe consentire all'AI generativa di comprendere le esigenze e le preferenze degli utenti, generando contenuti personalizzati e mirati.

2. **Apprendimento Continuo e Adattamento:** l'AGI, con la sua capacità di apprendere continuamente e di adattarsi a

nuove situazioni, potrebbe consentire all'AI generativa di:

- **Migliorare le proprie capacità nel tempo:** potrebbe aiutare l'AI generativa ad apprendere dai propri errori, a perfezionare le proprie tecniche e a generare contenuti sempre più sofisticati,
- **Generalizzare a nuovi domini e compiti:** potrebbe consentire all'AI generativa di applicare le proprie conoscenze e competenze a nuovi ambiti, generando contenuti in diverse forme e stili,
- **Collaborare con gli esseri umani:** potrebbe facilitare la collaborazione tra umani e AI generativa, consentendo agli utenti di guidare e influenzare il processo creativo.

3. **Ragionamento e Problem Solving:** L'AGI, con la sua capacità di ragionare, risolvere problemi e prendere decisioni, potrebbe consentire all'AI generativa di:

- **Generare contenuti più complessi e articolati:** potrebbe aiutare l'AI generativa a creare storie, articoli, programmi e altri contenuti che richiedono una struttura logica e una pianificazione complessa,
- **Risolvere problemi creativi:** potrebbe consentire all'AI generativa di superare blocchi creativi, esplorare soluzioni alternative e trovare nuove idee,
- **Automatizzare compiti complessi:** potrebbe consentire all'AI generativa di automatizzare la creazione di contenuti in diversi ambiti, liberando tempo e risorse per gli esseri umani.

L'AGI ha il potenziale per rivoluzionare l'AI generativa, trasformandola da uno strumento per la creazione di contenuti semplici a un partner creativo in grado di collaborare con gli esseri umani per raggiungere nuovi livelli di innovazione. L'integrazione dell'AGI nell'AI generativa potrebbe portare a una nuova era di creatività, in cui le macchine e gli esseri

umani lavorano insieme per esplorare le infinite possibilità dell'immaginazione.

Lo sviluppo dell'AGI solleva importanti questioni etiche:

- **Controllo e sicurezza**: come garantire che i sistemi di AGI siano sicuri e controllati, evitando comportamenti dannosi o imprevisti,
- **Impatto sociale**: come gestire l'impatto dell'AGI sul lavoro, l'economia e la società nel suo complesso,
- **Diritti e responsabilità**: quali diritti e responsabilità dovrebbero essere attribuiti ai sistemi di AGI.

In tal senso, la ricerca sull'AGI richiede un approccio multidisciplinare, che integri conoscenze provenienti da diversi campi, come l'informatica, la neuroscienza, la psicologia e la filosofia. È fondamentale affrontare le sfide scientifiche ed etiche associate all'AGI per garantire che questa tecnologia sia sviluppata e utilizzata in modo responsabile e benefico per l'umanità[6].

L'articolo ha esplorato le frontiere dell'Intelligenza Artificiale Generativa, evidenziando l'importanza di Explainable AI e Hybrid AI nel renderla più trasparente, etica e affidabile. Nel prossimo articolo approfondiremo ulteriori argomenti come il tema dell'"agente intelligente" nell'ambito dell'AI generativa e parleremo anche di AI neuro simbolica e di ingegneria dell'AI. Per saperne di più è possibile scaricare il nostro white paper completo "Generative Artificial Intelligence: Punti di forza, rischi e contromisure".

Note bibliografiche

[1] Pfeifer, R., & Bongard, J. (2006). How the body shapes the way we think: A new view of intelligence. MIT press.

[2][Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous

development of mental capabilities in computational agents. IEEE Transactions on Evolutionary Computation, 11(2), 151-180.

[3] Ziemke, T. (2001). Are robots embodied?. In Embodied artificial intelligence (pp. 75-96). Springer, Berlin, Heidelberg.

[4] Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. Journal of Artificial General Intelligence, 5(1), 1-48.

[5] Marcus, G. (2018). Deep Learning: A Critical Appraisal. arXiv preprint arXiv:1801.00631.

[6] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson Education Limited.

Profilo Autore

È laureato in Ingegneria Informatica ed in Sicurezza Informatica presso le Università di Roma La Sapienza e di Milano. Ha indirizzato la sua formazione nei settori della Cyber Security e Digital Forensics ottenendo i diplomi di perfezionamento in Data Protection e Data Governance; Criminalità Informatica e Investigazioni Digitali e Big Data, Artificial Intelligence.

Ha, altresì, conseguito l'Advanced Cybersecurity Graduate Certificate alla School of Engineering della Stanford University; Professional Certificates in Information Security; Incident Response Process; Digital Forensics e Cybersecurity Engineering and Software Assurance presso il Software Engineering Institute della Carnegie Mellon University.

Dal 1992 è nei ruoli del Ministero dell'Interno ove ricopre lincarico di Funzionario alla Sicurezza CIS. In tale veste contribuisce alla valutazione dei rischi cyber, all'implementazione delle misure di sicurezza e la risoluzione di incidenti informatici. Inoltre, offre consulenza tecnica

nel campo della Digital Forensics per l’Autorità giudiziaria, la Polizia giudiziaria e gli Studi legali.

Dal 2017 è Professore a contratto di Tecnologie per la Sicurezza Informatica presso alcune Università ove sviluppa le tematiche di Attack and Defense Strategies quali il penetration testing, la risk analysis, l’information security assessment, l’incident response e la digital forensics. Infine, è Autore di alcuni articoli e saggi sui temi della Sicurezza Informatica e dell’Informatica Giuridica consultabili su https://www.vincenzocalabro.it

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/articoli/nuovi-modelli-di-ia/